Cell, Vol. 113, 577–586, May 30, 2003, Copyright ©2003 by Cell Press

Riboswitches Control Fundamental Biochemical Pathways in *Bacillus subtilis* and Other Bacteria

Maumita Mandal,[1] Benjamin Boese,[2] Jeffrey E. Barrick,[3] Wade C. Winkler,[1] and Ronald R. Breaker[1,*]

[1]Department of Molecular, Cellular, and Developmental Biology

[2]Department of Chemical Engineering

[3]Department of Molecular Biophysics and Biochemistry

Yale University

New Haven, Connecticut 06520

Summary

Riboswitches are metabolite binding domains within certain messenger RNAs that serve as precision sensors for their corresponding targets. Allosteric rearrangement of mRNA structure is mediated by ligand binding, and this results in modulation of gene expression. We have identified a class of riboswitches that selectively recognizes guanine and becomes saturated at concentrations as low as 5 nM. In *Bacillus subtilis*, this mRNA motif is located on at least five separate transcriptional units that together encode 17 genes that are mostly involved in purine transport and purine nucleotide biosynthesis. Our findings provide further examples of mRNAs that sense metabolites and that control gene expression without the need for protein factors. Furthermore, it is now apparent that riboswitches contribute to the regulation of numerous fundamental metabolic pathways in certain bacteria.

Introduction

It is widely understood that the interplay of protein factors and nucleic acids guide the complex regulatory networks for genetic expression in modern cells. In most instances, protein factors appear to be well-suited agents for maintaining genetic expression networks. Proteins can adopt complex shapes and carry out a variety of functions that permit living systems to sense accurately their chemical and physical environments. Protein factors that respond to metabolites typically act by binding DNA to modulate transcription initiation (e.g., the Lac repressor protein; Matthews and Nichols, 1998) or by binding RNA to control either transcription termination (e.g., the PyrR protein; Switzer et al., 1999) or translation (e.g., the TRAP protein; Babitzke and Gollnick, 2001). Protein factors respond to environmental stimuli by various mechanisms such as allosteric modulation or posttranslational modification, and are adept at exploiting these mechanisms to serve as highly responsive genetic switches (e.g. see Ptashne and Gann, 2002).

In addition to the widespread participation of protein factors in genetic control, it is also known that RNA can take an active role in genetic regulation. Recent studies have begun to reveal the substantial role that small noncoding RNAs play in selectively targeting mRNAs for destruction, which results in downregulation of gene expression (see Hannon, 2002 and references therein). This process of RNA interference takes advantage of the ability of short RNAs to recognize the intended mRNA target selectively via Watson-Crick base complementation, after which the bound mRNAs are destroyed by the action of proteins. RNAs are ideal agents for molecular recognition in this system because it is far easier to generate new target-specific RNA factors through evolutionary processes than it would be to generate protein factors with novel but highly specific RNA binding sites.

Many studies have now confirmed that the complex three-dimensional shapes that some RNA molecules adopt can mimic protein receptors and antibodies in their ability to selectively bind proteins or even small molecules (Gold et al., 1995; Hermann and Patel, 2000). Furthermore, RNAs exhibit sufficient structural complexity to permit the formation of allosteric domains that undergo structural and functional modulation upon ligand binding (Soukup and Breaker, 1999a; Seetharaman et al., 2001). Natural RNAs also are capable of binding nucleotides, as demonstrated by the group I self-splicing RNA, which binds guanosine or its phosphorylated derivatives (McConnell et al., 1993). More recently, evidence has been provided which indicates that direct binding of ATP by an RNA is essential for packaging DNA into a viral capsid (Shu and Guo, 2003). Recognition of this potential for more sophisticated RNA function has fostered speculation that some mRNAs might bind metabolites directly, and that this binding could lead to genetic modulation (e.g. see Gold et al., 1997; Stormo and Ji, 2001). Indeed, we (Nahvi et al., 2002; Winkler et al., 2002a, 2002b, W.C. Winkler et al., submitted) and others (Mironov et al., 2002; McDaniel et al., 2003; Epshtein et al., 2003) have now confirmed that highly structured domains residing within noncoding regions of certain bacterial mRNAs serve as metabolite-responsive genetic switches.

The known riboswitches bind their target metabolites with high affinity and precision, which are essential characteristics for any type of molecular switch that can permit accurate and sensitive genetic control. For example, a recently identified riboswitch that responds to the coenzyme *S*-adenosylmethionine (SAM) binds its target with a dissociation constant (K_d) of ~4 nM (W.C. Winkler et al., submitted). Furthermore, the riboswitch can discriminate ~100-fold against *S*-adenosylhomocysteine, which is a natural metabolite that differs from SAM by a single methyl group and an associated positive charge. We have speculated that genetic control involving riboswitches is a widespread phenomenon with regard to its biological distribution and the target molecules that are monitored (Nahvi et al., 2002). The observations that certain mRNAs from archaeal organisms carry riboswitch-like domains (Stormo and Ji, 2001; Rodionov et al., 2002) and that several mRNAs from fungi and plants bind thiamine pyrophosphate (TPP) (Sudarsan et al., 2003) support this hypothesis.

*Correspondence: ronald.breaker@yale.edu

We have continued to examine additional candidate mRNAs for the possible participation of riboswitches in genetic control. We were intrigued by the genetic regulation of purine transport and purine biosynthesis pathways in bacteria, which are fundamental to the metabolic maintenance of nucleotides and nucleic acids (Switzer et al., 2002). In *B. subtilis*, numerous genes are involved in the biosynthesis of purines (*pur* operon with 12 genes; Ebbole and Zalkin, 1987) and in the salvage of purine bases from degraded nucleic acids. A regulatory protein factor has been proposed to participate in the control of the *xpt-pbuX* operon that encodes a xanthine phosphoribosyltransferase and a xanthine-specific purine permease, respectively (Christiansen et al., 1997). However, although the PurR protein is known to serve as a repressor of transcription in the presence of elevated adenine concentrations (Weng et al., 1995), no protein with corresponding function that responds to guanine has been identified in *B. subtilis*.

Herein, we report that the *xpt-pbuX* operon is controlled by a riboswitch that exhibits high affinity and high selectivity for guanine. This class of riboswitches is present in the 5′-untranslated region (5′-UTR) of five transcriptional units in *B. subtilis*, including that of the 12-gene *pur* operon. Thus, we conclude that direct binding of guanine by mRNAs serves as a critical determinant of metabolic homeostasis for purine metabolism in certain bacteria. Furthermore, we have determined that the known classes of riboswitches, which respond to seven distinct target molecules, appear to control at least 68 genes in *B. subtilis* that are of fundamental importance to central metabolic pathways. These findings indicate that riboswitches play a substantial role in metabolic regulation in living systems, and that direct interaction between small metabolites and RNA is a significant and widespread form of genetic regulation in bacteria.

Results and Discussion

A Conserved Domain in the 5′-UTR of Several *B. subtilis* mRNAs

Recent demonstrations that genetic control can be mediated by direct contacts between metabolites and mRNAs prompted us to examine the regulatory mechanisms of purine metabolism for possible riboswitch involvement. The *xpt-pbuX* operon is regulated by guanine, hypoxanthine, and xanthine. These purine compounds share chemical similarity and are adjacent to each other in the pathways of purine salvage. In contrast to the *pur* operon, regulation of the *xpt-pbuX* operon remains unaffected by adenine in a strain in which adenine deaminase is inactive (Christiansen et al., 1997). These observations had fostered speculation that an unidentified protein factor might be involved in guanine recognition (Ebbole and Zalkin, 1987); however, such a genetic factor has not been identified. Moreover, we noted that the 5′-UTR of the *xpt-pbuX* mRNA is rather large (185 nucleotides), and could be sufficient to accommodate a riboswitch domain.

Riboswitches are typically composed of two functional domains: an aptamer that selectively binds its target metabolite and an expression platform that responds to metabolite binding and controls gene expression by allosteric means. The most conserved portion of known riboswitches is the aptamer domain, whereas the adjoining expression platform can vary widely in both sequence and secondary structure. The high sequence conservation of the aptamer is due to the fact that the RNA must retain its ability to form a receptor for a chemical that does not change through evolution. In contrast, the expression platform can form one of a great diversity of structures that permit genetic control in response to ligand binding by the aptamer domain. We exploited this evolutionary conservation to conduct a database search for *xpt-pbuX* 5′-UTR sequences that are present in other *B. subtilis* genes and also in other bacterial species. We identified five transcriptional units within *B. subtilis* that closely correspond in sequence and predicted secondary structure with nucleotides 14 through 82 of the *xpt-pbuX* 5′-UTR (Figure 1). A total of 32 representatives of this domain were identified amongst several gram-positive and gram-negative bacteria, although we speculate that additional representatives could be found by conducting more sophisticated database searches.

From this representative set of RNAs, we identified a consensus sequence and secondary structure for the conserved RNA motif termed the "G box" (Figure 2A). The secondary structure of the G box is composed of a three-stem (P1 through P3) junction, wherein significant sequence conservation occurs within P1 and in the unpaired regions. Furthermore, we find that stems P2 and P3 both favor seven base pairs in length with one- or two-base mismatches permitted. This unusual conservation of stem length implies that these structural elements establish distance and orientation constraints of their stem-loop sequences relative to the three-stem junction. Some base-pairing potential exists between the two stem-loop sequences, which might permit the formation of a pseudoknot. These characteristics indicate that G-box domains most likely use conserved secondary- and tertiary-structure elements to adopt a precise three-dimensional fold.

The G Box RNA from the *xpt-pbuX* 5′-UTR of *B. subtilis* Binds Guanine

We prepared two RNA constructs based on the *xpt-pbuX* 5′-UTR of *B. subtilis* to examine whether the mRNA selectively binds guanine or its closest analogs. A double-stranded DNA template corresponding to the entire 5′-UTR and the first four codons of the *xpt-pbuX* mRNA was generated by PCR using primers that introduced a promoter sequence for T7 RNA polymerase and several nucleotide additions and mutations that permit further manipulation (Figure 2B; see Experimental Procedures). A truncated form of this construct also was created by PCR that encompasses the 5′ half of the UTR. Upon transcription, the shorter DNA template generates a 93-nucleotide transcript termed 93 *xpt*, while the longer template produces a 201-nucleotide transcript termed 201 *xpt*.

These precursor RNAs were 5′ ^{32}P-labeled and subjected to an in-line probing assay (e.g. see Soukup and Breaker, 1999b; Nahvi et al., 2002) wherein the spontaneous cleavage of RNA linkages within an aptamer is monitored in the presence and absence of its corre-

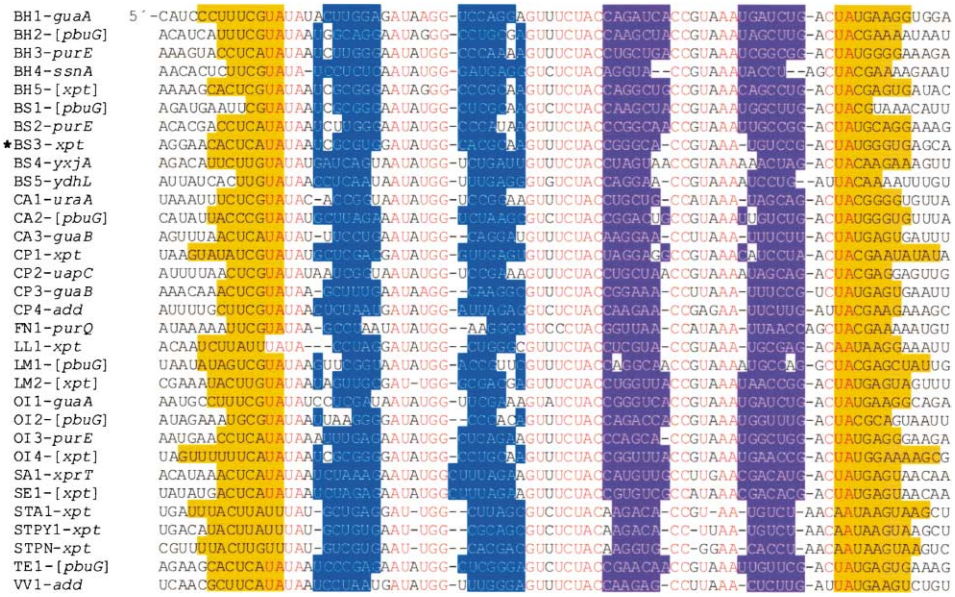

Figure 1. A Highly Conserved Domain Is Present in the 5′-UTR of Certain Gram-Positive and Gram-Negative Bacterial mRNAs

Depicted is an alignment of 32 representative mRNA domains from bacteria that conform to the G box consensus sequence. Regions shaded orange, blue, and purple identify base-pairing potential of stems P1, P2, and P3, respectively. Nucleotides in red are conserved in greater than 90% of the examples. The asterisk identifies the representative (*xpt-pbuX* 5′-UTR) that was examined in this study. It is important to note that three representatives (BS5, CP4, and VV1) that carry a C to U mutation in the conserved core (in the P3-P1 junction) appear to be adenine-specific riboswitches (M.M. and R.R.B., unpublished observations). Gene names are as annotated in GenBank, the SubtiList database, or based on protein similarity searches (brackets). Organisms abbreviations are as follows: *Bacillus halodurans* (BH), *Bacillus subtilis* (BS), *Clostridium acetobutylicum* (CA), *Clostridium perfringens* (CP), *Fusobacterium nucleatum* (FN), *Lactococcus lactis* (LL), *Listeria monocytogenes* (LM), *Oceanobacillus iheyensis* (OI), *Staphylococcus aureus* (SA), *Staphylococcus epidermidis* (SE), *Streptococcus agalactiae* (STA), *Streptococcus pyogenes* (STPY), *Streptococcus pneumoniae* (STPN), *Thermoanaerobacter tengcongensis* (TE), and *Vibrio vulnificus* (VV).

sponding ligand. We find that the patterns of spontaneous cleavage of the 93 *xpt* (Figure 2C) and the 201 *xpt* (Figure 3A) RNAs undergo significant alteration upon addition of guanine at a concentration of 1 μM. Both hypoxanthine and xanthine also induce modulation of spontaneous cleavage at this concentration. Specifically, four major regions exhibit ligand-mediated reduction in spontaneous cleavage (Figures 2B and 2C). How-

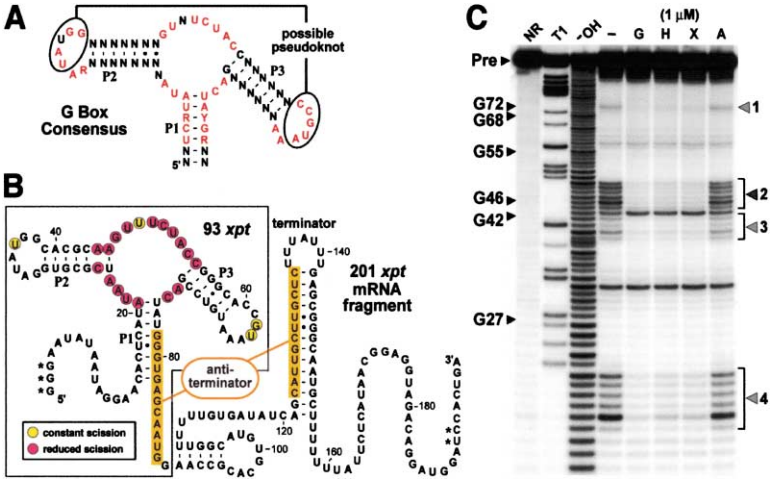

Figure 2. The G Box RNA of the *xpt-pbuX* mRNA in *B. subtilis* Responds Allosterically to Guanine

(A) The consensus sequence and secondary model for the G box RNA domain that resides in the 5′-UTR of genes that are largely involved in purine metabolism. Phylogenetic analysis is consistent with the formation of a three-stem (P1 through P3) junction. Nucleotides depicted in red and black are present in greater than 90% and 80% of the representatives examined (Figure 2). Encircled nucleotides exhibit base complementation, which might indicate the formation of a pseudoknot.

(B) Sequence- and ligand-induced structural alterations of the 5′-UTR of the *xpt-pbuX* transcriptional unit. The putative antiterminator interaction is highlighted in orange. Nucleotides that undergo structural alteration as determined by in-line probing (from C) are identified with red circles. The 93 *xpt* fragment (boxed) of the 201 *xpt* RNA retains guanine binding function. Asterisks denote alterations to the RNA sequence that facilitate in vitro transcription (5′ terminus) or that generate a restriction site (3′ terminus). Nucleotide numbers begin at the first nucleotide of the natural transcription start site. The translation start codon begins at position 186.

(C) Guanine and related purines selectively induce structural modulation of the 93 *xpt* mRNA fragment. Precursor RNAs (Pre; 5′ ^{32}P-labeled) were subjected to in-line probing by incubation for 40 hr in the absence (–) or presence of guanine (G), hypoxanthine (H), xanthine (X), and adenine (A). Lanes designated NR, T1, and $^{-}$OH contain RNA that was not reacted, subjected to partial digestion with RNase T1 (G-specific cleavage), or subjected to partial alkaline digestion, respectively. Selected bands corresponding to G-specific cleavage are identified. Regions 1 through 4 identify major sites of ligand-induced modulation of spontaneous RNA cleavage.

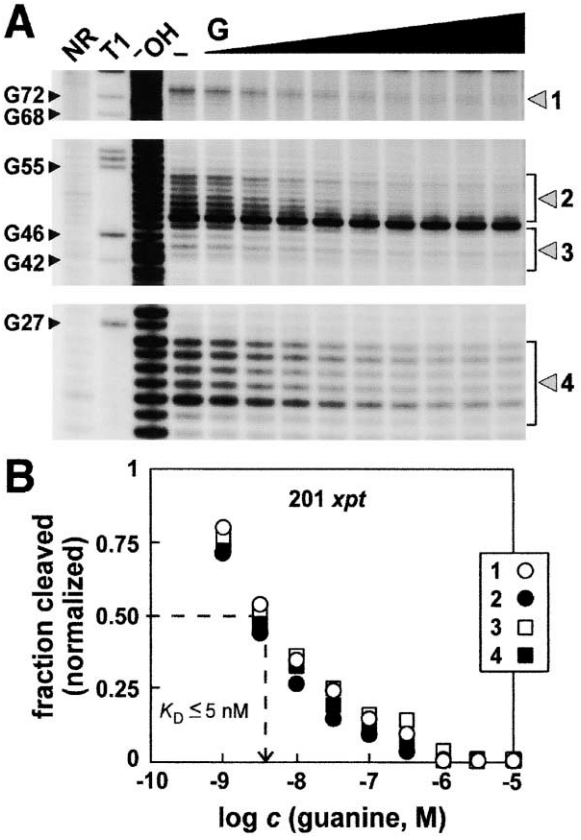

Figure 3. The 201 *xpt* mRNA Leader Binds Guanine with High Affinity

(A) In-line probing reveals that spontaneous RNA cleavage of the 201 *xpt* RNA at four regions decreases with increasing guanine concentrations. Only those locations of the PAGE image corresponding to the four regions of modulation as indicated in Figure 3C are depicted. Other details and notations are as described in the legend to Figure 3C.

(B) Plot depicting the normalized fraction of RNA that experienced spontaneous cleavage versus the concentration of guanine for modulated regions 1 through 4 in (A). Fraction cleaved values were normalized to the maximum cleavage measured in the absence of guanine and to the minimum cleavage measured in the presence of 10 μM guanine. The apparent K_d value (less than or equal to 5 nM) reflects the limits of detection for these assay conditions.

ever, the presence of 1 μM adenine (and as much as 1 mM; data not shown) does not alter the pattern of RNA cleavage products. These results indicate that the G box domain in the 5′-UTR of the *B. subtilis xpt-pbuX* mRNA serves as an aptamer for guanine and related purines, and that this aptamer undergoes significant structural modulation upon ligand binding. In the context of a riboswitch, this allosteric function could be harnessed by the mRNA to modulate structural elements that regulate gene expression.

In a preliminary assessment of the affinity that the guanine aptamer has for its target, we conducted in-line probing with 201 *xpt* in the presence of various concentrations of guanine. As expected, increasing concentrations provided progressively decreasing amounts of spontaneous cleavage at the four major sites of structural modulation (Figure 3A). We observed half-maximum levels of modulation when a concentration of ~5 nM guanine is used for in-line probing (Figure 3B). Although this implies that the K_d for 201 *xpt* under these conditions is ~5 nM, it is important to note that the actual value might be somewhat lower because of the limitations of the in-line probing assay (see Experimental Procedures). In addition, the K_d was determined under nonphysiological conditions (e.g., high Mg^{2+} ion concentration and elevated pH), and so the binding affinity might be somewhat different in vivo. However, using this number for comparison, the affinity of the 201 *xpt* RNA for guanine is more than 10,000-fold greater than that of the *Tetrahymena* group I ribozyme for its guanosine monophosphate substrate (McConnell et al., 1993). This difference most likely reflects the relative differences in concentrations of the two compounds that the RNAs experience inside their respective cellular environments.

The Guanine Aptamer Discriminates against Many Purine Analogs

To maintain precise metabolic homeostasis, the cell not only must be able to sense the concentration of its target metabolite, but also must prevent regulatory crosstalk with other compounds that otherwise might inadvertently trigger genetic modulation. Indeed, a hallmark of other riboswitches is the ability to discriminate between closely related metabolites. For example, the FMN and TPP riboswitches discriminate against the unphosphorylated coenzyme precursors thiamine and riboflavin by ~1000-fold (Winkler et al., 2002a, 2002b).

This requirement for obligate molecular discrimination against related metabolites is expected to be extreme with guanine riboswitches, as there are numerous purine nucleosides and nucleotides, purine bases, and purine-like compounds present in the cell. Using the in-line probing strategy described in Figure 3, we established the apparent K_d values of the 93 *xpt* RNA for a variety of purines and purine analogs. Hypoxanthine and xanthine exhibit K_d values that are closest to the value determined for guanine, while adenine has a K_d value in excess of 300 μM (Figure 4A). These results are consistent with the observation that adenine, unlike other purines, does not significantly repress expression of the *xpt-pbuX* operon (Christiansen et al., 1997). However, it is not clear whether hypoxanthine and xanthine might repress gene expression by directly binding a guanine riboswitch, or whether they might first be converted into guanine before influencing genetic control.

Interestingly, we find that alteration of every functionalized position on the guanine heterocycle causes a substantial loss of binding affinity (Figures 4B and 5). For example, the oxygen atom at position 6 of guanine is a significant determinant of molecular recognition, as demonstrated by the losses in apparent K_d for 2-aminopurine (>10,000-fold), 2-amino-6-bromopurine (1000-fold), and O^6-methylguanine (>100-fold). Most molecular interactions could be explained by invoking hydrogen-bonding contacts between the RNA and guanine with the exception of the molecular interaction at C8. Here, presumably the RNA structure creates a steric clash with analogs that carry additional bulk, such as 8-methylxanthine (>10,000-fold) and uric acid (>10,000-fold).

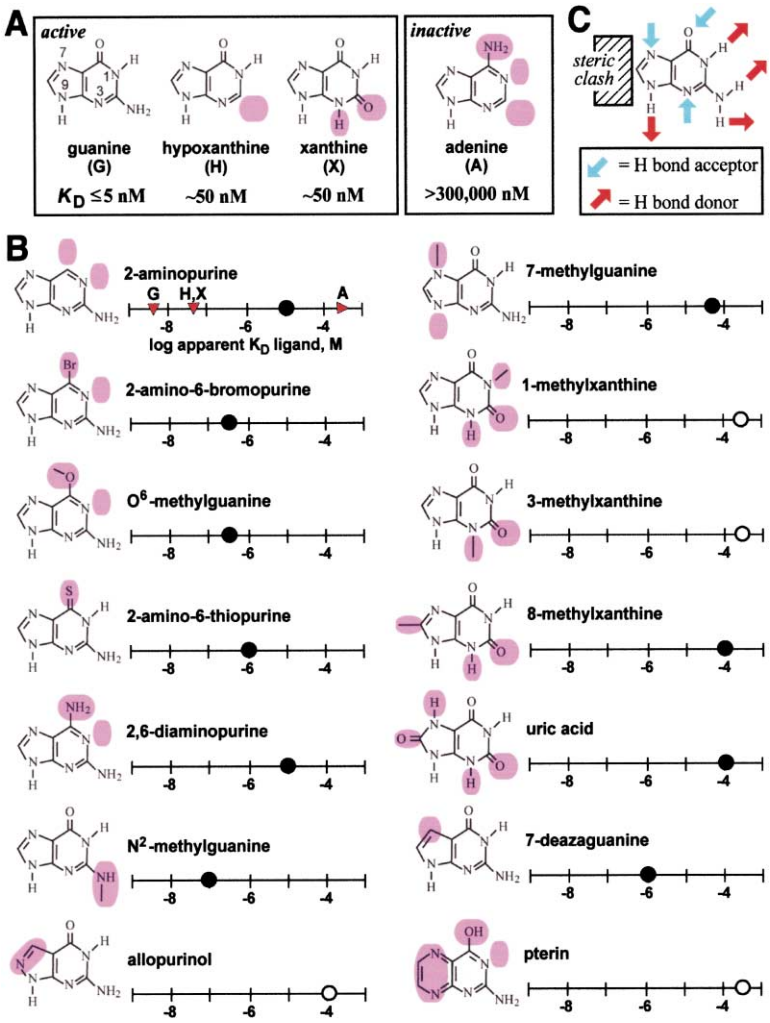

Figure 4. Molecular Discrimination by the Guanine-Binding Aptamer of the *xpt-pbuX* mRNA

(A) Chemical structures and apparent K_d values for guanine, hypoxanthine, and xanthine (active natural regulators of *xpt-pbuX* genetic expression in *B. subtilis*) versus that of adenine (inactive). Differences in chemical structure relative to guanine are shaded pink. K_d values were established as shown in Figure 4 with the 201 *xpt* RNA. Numbers on guanine represent the positions of the ring nitrogen atoms.

(B) Chemical structures and K_d values for various analogs of guanine reveal that all alterations of this purine cause a loss of binding affinity. Open circles identify K_d values that most likely are significantly higher than indicated, as concentrations of analog above 500 μM were not examined in this analysis. The apparent K_d values of G, H, X, and A as indicated are plotted as red triangles for comparison.

(C) Schematic representation of the molecular recognition features of the guanine aptamer in 201 *xpt*. Hydrogen bond formation at position 9 of guanine is expected because guanosine (K_d >100 μM) and inosine (K_d >100 μM), which are 9-ribosyl derivatives of guanine and hypoxanthine, respectively, do not exhibit measurable binding (see Figure 6).

A summary of the likely molecular recognition features that the guanine aptamer requires for maximum affinity is depicted in Figure 4C. However, we have not examined the likely possibility that significant binding affinity could be derived through base stacking. The presence of so many productive contacts between the RNA and all faces of guanine suggest that the ligand is most likely entirely engulfed by the aptamer's structure. This would also explain why the RNA is capable of generating recognition via steric occlusion of bulkier compounds such as uric acid. In certain biological environments, for example, uric acid can build up to high concentrations that permit crystallization. In such environments, a bacterium would require a high level of discrimination to prevent undesirable repression of guanine-regulated genes. In light of such molecular recognition challenges, it is not surprising that an RNA genetic switch would evolve extensive molecular contacts with its target compound.

Confirmation of Guanine Aptamer Function by Equilibrium Dialysis

We employed equilibrium dialysis to provide further evidence that the G box RNA from the *xpt-pbuX* operon binds guanine preferentially over other purines and purine analogs. A substantial shift in tritiated guanine is expected to occur in a two-chamber dialysis apparatus when an excess of functional RNA is added to one chamber (Figure 5A). Furthermore, this shifted equilibrium should return to unity upon addition of an excess of unlabeled competitor ligand. As expected, we observe that greater than 90% of tritiated guanine colocalizes with 93 *xpt* RNA, and subsequently redistributes when an excess of unlabeled guanine is introduced. In contrast, the presence of excess unlabeled analogs has no effect on colocalization of ^{3}H-guanine and the RNA (Figure 5B). Even the nucleoside guanosine (9-ribosylguanine) fails to restore equal distribution of guanine between the two chambers, which is consistent with our hypothesis that the RNA folds to form a tight pocket for the base alone.

Both in-line probing and equilibrium dialysis data indicate that this natural aptamer binds guanine with high-affinity and specificity. In a previous study, in vitro evolution was used to isolate a purine binding aptamer from a pool of random-sequence RNAs (Kiga et al., 1998). This engineered aptamer exhibits a K_d of 1.3 μM for guanine and shows only a 2- to 3-fold discrimination against hypoxanthine and xanthine. The lower specificity and affinity of this aptamer for selected purines is

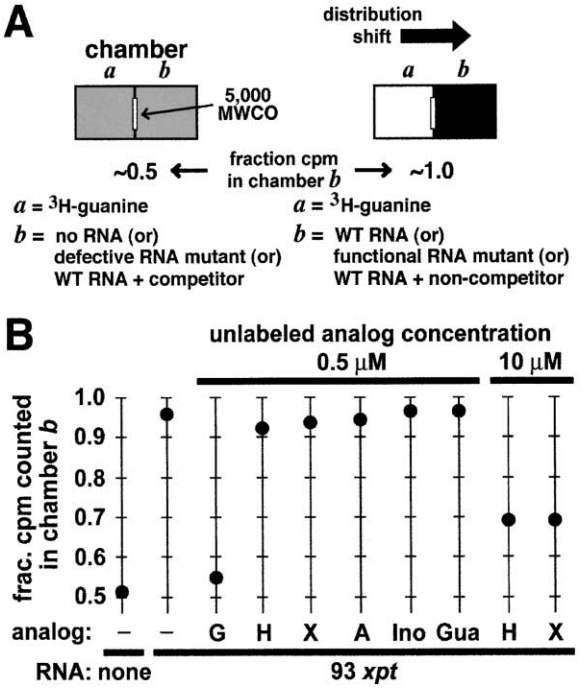

Figure 5. Confirmation of Guanine Binding Specificity by Equilibrium Dialysis

(A) An equilibrium dialysis strategy was used to confirm that in vitro-transcribed 93 *xpt* RNAs bind to guanine and can discriminate against various analogs. Each data point was generated by adding ^{3}H-guanine to chamber *a*, which is separated from RNA and other analogs by a dialysis membrane with a molecular weight cut off (MWCO) of 5,000 daltons. Left: If no guanine binding sites are present in chamber *b*, or if an excess of unlabeled competitor is present, then no shift in the distribution of tritium is expected. Right: If an excess of guanine binding RNAs are present in chamber *b*, and if no competitor is present, then a substantial shift in the distribution of tritium toward chamber *b* is expected.

(B) The 93 *xpt* RNA can shift the distribution of ^{3}H-guanine in an equilibrium dialysis apparatus, while analogs of guanine are poor competitors. The plot depicts the fraction of counts per minute (cpm) of tritium in chamber *b* relative to the total amount of cpm counted from both chambers. A value of ~0.5 is expected if no shift occurs, as is the case when RNA is absent (none), or in the presence of excess unlabeled competitor (G). A value approaching 1 is expected if the majority of ^{3}H-guanine is bound by the RNA in chamber *b* in the absence (–) of unlabeled analog, or in the presence of unlabeled analogs that do not serve as effective competitors under our assay conditions (100 nM ^{3}H-guanine, 300 nM RNA, and 500 nM analog). Ino and Gua represents inosine and guanosine, respectively.

due to the fact that only the N1, N7, and O6 positions are important for molecular recognition. In contrast, the G box RNA appears to make productive contacts with all available functional groups on guanine, presumably through hydrogen bonding (Figure 4C).

Aptamer Mutations Affect Guanine Binding and Genetic Control

A variety of mutations were introduced into the G box domain to examine the importance of several structural elements and conserved nucleotides (Figure 6A). The influence of these mutations on guanine binding was determined in the context of the 93 *xpt* RNA by using equilibrium dialysis. Mutations that independently disrupt the three stems (M1, M4, and M6) cause a loss of binding function, as does a variant RNA (M3) that carries two mutations in the central junction (Figure 6B). In contrast, the effects of the disruptive stem mutations are largely reversed by making compensatory mutations (M2, M5, and M7) that restore base pairing. These results are consistent with the phylogenetic analysis (Figure 1), which indicates that stem structure is important but that the precise sequence composition of these elements is of less importance.

Binding function of variant aptamers in vitro also correlates with genetic control in vivo. We have confirmed earlier findings that a reporter gene carrying the 5′-UTR of the *xpt-pbuX* mRNA is repressed by guanine, and to a lesser extent by hypoxanthine and xanthine (Christiansen et al., 1997). Specifically, transcriptional fusions were created between a β-galactosidase reporter gene and variant *xpt-pbuX* 5′-UTR sequences carrying the mutations described in Figure 6A. *B. subtilis* chromosomal transformants using the wild-type sequence exhibit the expected levels of genetic modulation (Figure 6C). Although the *xpt* aptamer exhibits dissociation constants for xanthine and hypoxanthine that are essentially identical in vitro, the differences in genetic modulation by these compounds in vivo might be due to differences in their cellular concentrations.

Aptamer variants with impaired guanine binding in vitro also exhibit a loss of β-galactosidase repression (Figure 6D). Furthermore, restoration of base pairing in stems P1 through P3 results in restored genetic control. The M2 variant is of particular interest because it not only exhibits restored genetic control, but also provides modest expression of β-galactosidase in the absence of guanine. Riboswitch function requires the action of an aptamer for molecular sensing as well as an expression platform that transduces RNA-ligand complex formation into a genetic response. Examples of TPP and FMN riboswitches (Winkler et al., 2002a, 2002b; Mironov et al., 2002) appear to function by differential formation of terminator and anti-terminator structures. Such ligand-induced formation of transcription antitermination structures also appears to be the basis of expression platform mechanisms used by numerous SAM riboswitches (W.C. Winkler et al., submitted; McDaniel et al., 2003). Construct M2 carries three mutations within the putative antiterminator structure of the *xpt-pbuX* leader, and thus is expected to exhibit an overall reduction of reporter gene expression because these mutations should bias structure folding toward terminator stem formation.

The results of these mutational and functional analyses confirm the major features of the secondary structure model (P1 through P3) and demonstrate that they are critical for metabolite binding. Furthermore, the correlation between ligand binding and genetic control indicates that the G box and adjacent nucleotides of the *xpt-pbuX* leader sequence operate in concert to function as a guanine-dependent riboswitch, most likely by operating via allosteric control of transcription termination.

Riboswitches Control Fundamental Biochemical Pathways

Our findings indicate that the G box RNA of the *xpt-pbuX* operon is a key structural element of a guanine-

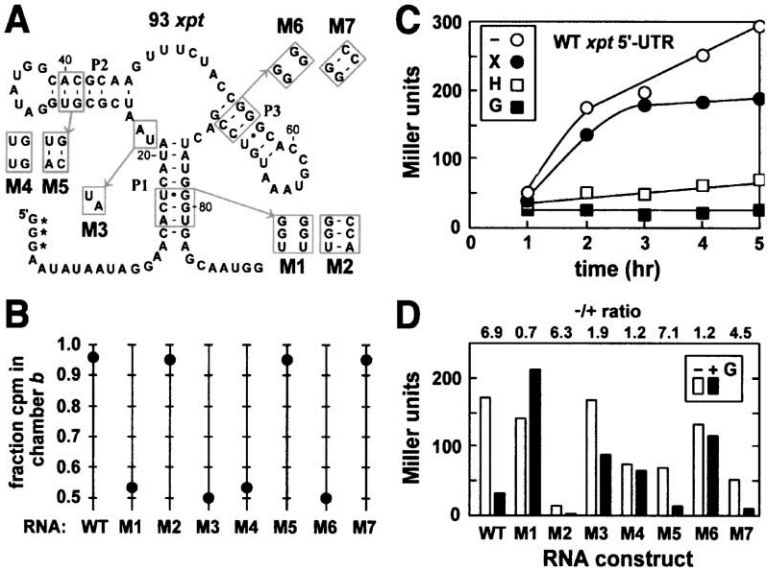

Figure 6. The Binding and Genetic Control Functions of Variant Guanine Riboswitches

(A) Mutations used to examine the importance of various structural features of the guanine aptamer domain.

(B) Examination of the binding function of aptamer variants by equilibrium dialysis. WT designates the wild-type 93 *xpt* construct. Details are as described for Figure 5.

(C) Genetic modulation of a β-galactosidase reporter gene upon the introduction of various purines as indicated.

(D) Regulation of β-galactosidase reporter gene expression by WT and mutants M1 through M7. Open and filled bars represent enzyme activity generated when growing cells in the absence and presence of guanine, respectively.

sensing riboswitch that exhibits extraordinary affinity and selectivity for its target. In *B. subtilis*, this general riboswitch motif appears to control at least five transcriptional units (Figure 1). Although the precise function of several of the gene products in this regulon have not been clearly defined, the known genes from *B. subtilis* and from other organisms are mostly related to purine metabolism. In the current study, we have confirmed that a guanine riboswitch controls the *xpt-pbuX* operon, which encodes for purine salvage proteins involved in the import and phosphoribosylation of xanthine. Interestingly, a guanine riboswitch probably provides a regulatory contribution to the 12-gene *purEKBCSQLFMNHD* operon for de novo purine biosynthesis. Transcription initiation of this large biosynthetic operon is controlled by the PurR repressor protein, which causes reduced transcription initiation when adenine or adenosine is added to cell cultures (Switzer et al., 2002). However, it is also known that this transcriptional unit is prematurely terminated in the presence of guanine (Ebbole and Zalkin, 1987; Switzer et al., 2002). Based on our results, we propose that the G box domain within the 5′-UTR of this large *pur* operon is responsible for guanine-dependent riboswitch regulation, and we suggest that the genetic regulatory mechanism might be similar to that proposed herein for the *xpt-pbuX* operon.

The distribution of G box domains in *B. subtilis* and other bacteria suggests that this class of metabolite binding RNAs controls a regulon that is essential for cell survival. In *B. subtilis*, guanine riboswitches (or related adenine-dependent riboswitches; see the legend to Figure 1) appear to provide at least some contribution to the genetic regulation of 17 genes. The discovery of guanine-dependent riboswitches adds to a growing list of similar metabolite-sensing RNAs (W.C. Winkler et al., submitted). For example, a class of riboswitches that responds to SAM (McDaniel et al., 2003; Epshtein et al., 2003; W.C. Winkler et al., submitted) controls a regulon of as many as 26 genes that are involved in coenzyme biosynthesis, amino acid metabolism, and sulfur metabolism. When included with genes that are controlled by other riboswitch classes, at least 68 genes (nearly 2% of its total genetic complement) are under riboswitch control (Figure 7).

Riboswitches for ligands such as guanine and SAM apparently are serving as master control molecules whose concentrations are being monitored to ensure homeostasis of a much wider set of metabolic pathways. Riboswitches also seem to permit metabolite surveillance and genetic control with the same level of precision and efficiency as that exhibited by protein factors. Therefore, these RNA switches could have emerged late in the evolution of modern biochemical architectures because they are functionally comparable to genetic switches made of protein. However, given their fundamental role in metabolic maintenance and the widespread phylogenetic distribution of certain riboswitches, we speculate that aptamer domains similar to these might have been the primary mechanism by which RNA-world organisms detected metabolites and controlled biochemical pathways before the emergence of proteins.

Conclusions

This demonstration that guanine is sensed by metabolite binding mRNAs expands the known classes of riboswitches and provides additional evidence that certain bacterial RNAs are responsible for monitoring the concentrations of critical coenzymes and other compounds that are fundamental to all living systems. Phylogenetic analyses and biochemical data indicate that many bacteria and, in some instances, eukaryotes (Sudarsan et al., 2003) entrust riboswitches to sense essential metabolites and mediate genetic control. Although protein factors undoubtedly could be used to carry out these important regulatory tasks, we conclude that highly structured RNAs are well suited for this role. If RNA polymers were a poorly suited medium for generating metabolite receptors with high affinity and precision, then we would expect that evolution would have long ago replaced them by protein factors.

We have speculated (e.g., see Nahvi et al., 2002; Wink-

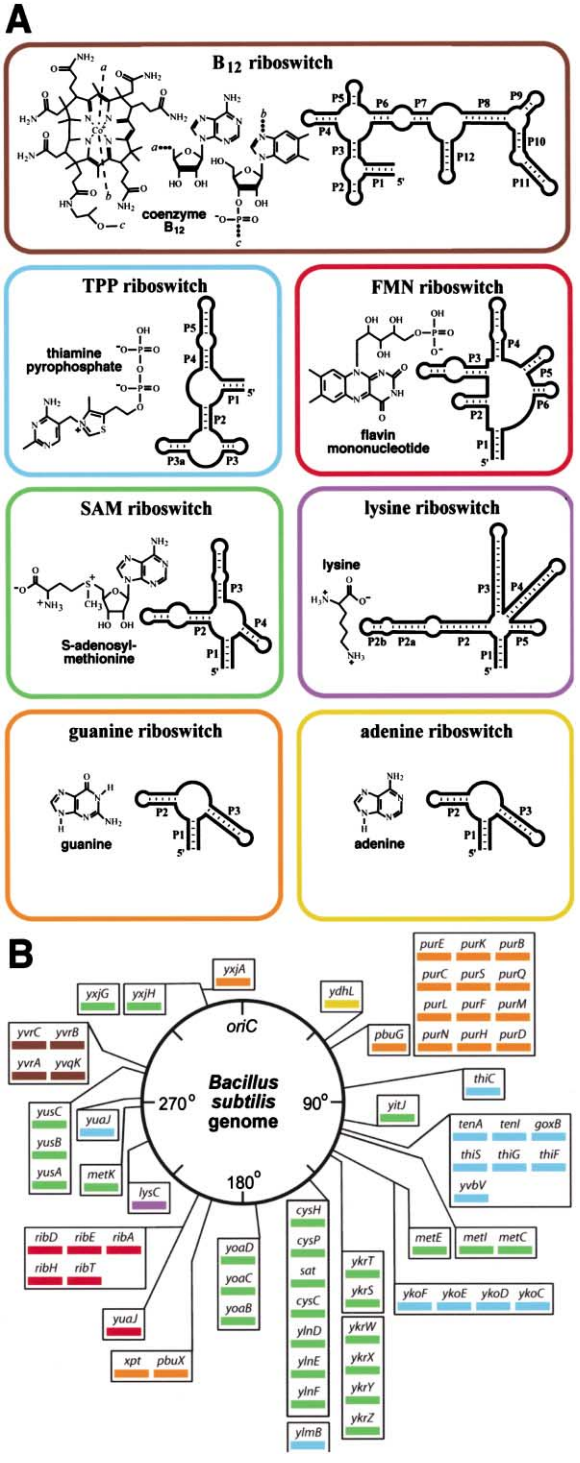

Figure 7. Riboswitches Participate in Fundamental Genetic Control

(A) Schematic representations of the seven known riboswitches and the metabolites they sense. The secondary structure models were obtained as follows: coenzyme B_{12} (revised from Nahvi et al., 2002; Ali Nahvi, J.E.B., and R.R.B., unpublished data); TPP (Winkler et al., 2002a; Sudarsan et al., 2003; Rodionov et al., 2002); FMN (Winkler et al., 2002b; Vitreschak et al., 2002); SAM (Grundy and Henkin, 1998; W.C. Winkler et al., submitted); guanine (this work); lysine (Narasimhan Sudarsan, John K. Wickiser, and R.R.B., unpublished data); and adenine (this work and M.M. and R.R.B., unpublished data). Coenzyme B_{12} is depicted in exploded form wherein *a*, *b*, and *c* designate covalent attachment sites between fragments.

(B) Genetic map of *B. subtilis* riboswitch regulons and their positions on the bacterial chromosome. Genes are controlled by riboswitches as identified by matching color. All nomenclature is derived from the SubtiList database release R16.1 (Moszer et al., 1995) except for *metI* and *metC*, which are recent designations (Auger et al., 2002).

ler et al., 2002a) that riboswitches might be derivatives of an ancient genetic control system that monitored metabolic and environmental signals before the evolutionary emergence of proteins. Interestingly, we find that each of the metabolite targets of riboswitches has been proposed to come from an RNA world (White, 1976; Benner et al., 1989; Jeffares et al., 1998; Jadhav and Yarus, 2002). The identification of guanine as a trigger for riboswitches is consistent with our hypothesis that metabolite sensing RNAs might have originated very early in evolution. We have recently identified another class of riboswitches that responds to the amino acid lysine (Figure 7). Although all riboswitches could be more recent evolutionary inventions, even the origin of the lysine riboswitch might date from before the last common ancestor and back to a time when living systems were transitioning from a pure RNA world to a more modern metabolic state that made use of encoded protein synthesis.

Experimental Procedures

Chemicals and Oligonucleotides

Guanine and its analogs xanthine, hypoxanthine, adenine, guanosine, 7-methylguanine, N^2-methylguanine, 1-methylxanthine, 3-methylxanthine, 8-methylxanthine, 2-aminopurine, 2,6-diaminopurine, allopurinol, 2-amino-6-mercaptopurine, lumazine, and guanine-8-^{3}H hydrochloride were purchased from Sigma. Inosine, uric acid, 2-amino-6-bromopurine, O^6-methylguanine, and pterin were purchased from Aldrich.

DNA oligonucleotides were synthesized by the Keck Foundation Biotechnology Resource Center at Yale University, purified by denaturing PAGE, and eluted from the gel by crush-soaking in 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl, and 1 mM EDTA. Oligonucleotides were recovered from solution by precipitation with ethanol.

Phylogenetic Analyses

G box domains were identified by sequence similarity to the *xpt-pbuX* 5′-UTR by conducting a BLASTN search of GenBank using default parameters. These hits were expanded by searching for degenerate matches to the pattern (<<<< [2] TA [6] <<< [2] ATNNGG [2] >>> [5] GTNTCTAC [3] <<<<< [3] CCNNNAA [3] >>>>> [5] >>>>) using the program SequenceSniffer (J.E.B. and R.R.B., unpublished algorithm). Angled brackets indicate base pairing. Bracketed numbers are variable gaps with constrained maximum lengths denoted. A total of four violations of this pattern were permitted when forming the phylogeny depicted in Figure 1. It is important in this instance to note that only the BS3-xpt domain (that of the *xpt-pbuX* leader) has been shown to bind guanine. We have demonstrated that the molecular specificity of the VV1 representative is for adenine and not guanine (M.M. and R.R.B., unpublished data). Given the possible trivial means by which a guanine binding RNA aptamer might be altered to bind adenine (e.g., a C to U change if the C residue is used by the aptamer to make a Watson-Crick-pairing interaction with guanine), we cannot rule out the possibility that other representatives also have altered molecular recognition.

In-Line Probing of RNA Constructs

The *B. subtilis* 201 *xpt* leader and truncated 93 *xpt* aptamer RNAs were prepared by in vitro transcription using T7 RNA polymerase

and the appropriate PCR DNA templates, and were subsequently 5′ ^{32}P-labeled using a protocol similar to that described previously (Seetharaman et al., 2001). Labeled precursor RNAs (2 nM) were subjected to in-line probing using conditions similar to those described previously (Winkler et al., 2002b). Reactions (10 μl) were incubated for 40 hr at 25°C in a buffer containing 50 mM Tris (pH 8.5 at 25°C), 20 mM $MgCl_2$, and 100 mM KCl in the presence or absence of purines as indicated for each experiment. Purine concentrations ranging from 1 nM to 10 μM were typically employed but ranged as high as 300 μM for poor binding ligands. Denaturing 10% PAGE was used to separate spontaneous cleavage products and a Molecular Dynamics PhosphorImager was used to view the results. Quantitation of spontaneous cleavage yields was achieved by using ImageQuaNT software. Since concentrations of RNA below 2 nM for in-line probing cannot be used easily due to insufficient levels of signal, apparent K_d values near this concentration reflect the maximum possible value.

Equilibrium Dialysis

Equilibrium dialysis assays were conducted using a DispoEquilibrium Dialyzer (ED-1, Harvard Bioscience), wherein chambers *a* and *b* were separated by a 5,000 MWCO membrane. The final composition of buffer included 50 mM Tris-HCl (pH 8.5 at 25°C), 20 mM $MgCl_2$, and 100 mM KCl (30 μl delivered to each chamber). Chamber *a* also contained 100 nM ^{3}H-guanine, while chamber *b* also contained 300 nM of *xpt* RNA constructs as indicated for each experiment. After 10 hr of equilibration at 25°C, a 5 μl aliquot from each chamber was removed for quantitation by liquid scintillation counter. When appropriate, an additional 5 μl of buffer was added to *a* and an equivalent volume of buffer containing 500 nM unlabeled purine was added to *b*. After an additional 10 hr incubation at 25°C, 5 μl aliquots were again drawn for quantitation of tritium distribution.

Construction of *xpt-lacZ* Fusions

Genetic manipulations were conducted using approaches similar to those described elsewhere (W.C. Winkler et al., submitted). Briefly, a DNA construct encompassing nt −121 to +197 relative to the transcription start site of the *xpt-pbuX* operon from *B. subtilis* strain 1A40 (Bacillus Genetic Stock Center, Columbus, OH) was PCR amplified as an EcoR1-BamH1 fragment. The product was cloned into pDG1661 at a site directly upstream of the *lacZ* reporter gene. Mutants were created within the engineered pDG1661 by using the appropriate primers and the QuickChange site-directed mutagenesis kit (Stratagene). Plasmid variants were integrated into the *amyE* locus of strain 1A40. Transformants were selected for chloramphenicol (5 μg/ml) resistance and screened for sensitivity to spectinomycin (100 μg/ml). The integrity of each construct was confirmed by sequencing.

Guanine-Mediated Modulation of β-Galactosidase Expression

B. subtilis cells were grown with shaking at 37°C in minimal media containing 0.4% w/v glucose, 20 g/l $(NH_4)_2SO_4$, 25 g/l $K_2HPO_4{\cdot}3H_2O$, 6 g/l KH_2PO_4, 1 g/l sodium citrate, 0.2 g/l $MgSO_4{\cdot}7H_2O$, 0.2% glutamate, 5 μg/ml chloramphenicol, 50 μg/ml L-tryptophan, 50 μg/ml L-lysine, and 50 μg/ml L-methionine. Purines were added at a final concentration of 0.5 mg/ml. Cells at mid exponential stage (A_{595} of 0.1) were harvested by centrifugation and resuspended in minimal media in the absence or presence of a purine (0.5 mg/ml) as indicated for each experiment. Although the poor solubility of guanine causes the formation of a detectable level of precipitate at this concentration, no adverse affects of cell growth were observed. Unless otherwise specified, cells were incubated for an additional 3 hr before performing β-galactosidase assays. Data presented in Figure 6C were generated as described above with the exception that β-galactosidase assays were performed at the times indicated.

Acknowledgments

We thank members of the Breaker laboratory for comments on the manuscript and Kausik Chakrabarti for helpful discussions. This work was supported by grants from the NIH (GM 559343) and the NSF (EIA-0129939). R.R.B. is also supported by a fellowship from the David and Lucile Packard Foundation.

Received: March 31, 2003
Revised: April 29, 2003
Accepted: April 29, 2003
Published: May 29, 2003

References

Auger, S., Yuen, W.H., Danchin, A., and Martin-Verstraete, I. (2002). The *metIC* operon involved in methionine biosynthesis in *Bacillus subtilis* is controlled by transcription antitermination. Microbiology *148*, 507–518.

Babitzke, P., and Gollnick, P. (2001). Posttranscriptional initiation control of tryptophan metabolism in *Bacillus subtilis* by the trp RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. J. Bacteriol. *183*, 5795–5802.

Benner, S.A., Ellington, A.D., and Tauer, A. (1989). Modern metabolism as a palimpsest of the RNA world. Proc. Natl. Acad. Sci. USA *86*, 7054–7058.

Christiansen, L.C., Schou, S., Nygaard, P., and Saxild, H.H. (1997). Xanthine metabolism in *Bacillus subtilis*: characterization of the *xpt-pbuX* operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. J. Bacteriol. *179*, 2540–2550.

Ebbole, D.J., and Zalkin, H. (1987). Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for de novo purine nucleotide synthesis. J. Biol. Chem. *262*, 8274–8287.

Epshtein, V., Mironov, A.S., and Nudler, E. (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. Proc. Natl. Acad. Sci. USA *100*, 5052–5056.

Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995). Diversity of oligonucleotide functions. Annu. Rev. Biochem. *64*, 763–797.

Gold, L., Brown, D., He, Y., Shtatland, T., Singer, B.S., and Wu, Y. (1997). From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops. Proc. Natl. Acad. Sci. USA *94*, 59–64.

Grundy, F.J., and Henkin, T.M. (1998). The S box regulon: a new global transcription termination control system for methionine and cystenine biosynthesis genes in gram-positive bacteria. Mol. Microbiol. *30*, 737–749.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

Hermann, T., and Patel, D. (2000). Adaptive recognition by nucleic acid aptamers. Science *287*, 820–825.

Jadhav, V.R., and Yarus, M. (2002). Coenzymes as coribozymes. Biochimie *84*, 877–888.

Jeffares, D.C., Poole, A.M., and Penny, D. (1998). Relics from the RNA world. J. Mol. Evol. *46*, 18–36.

Kiga, D., Futamura, Y., Sakamoto, K., and Yokoyama, S. (1998). An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. Nucleic Acids Res. *26*, 1755–1760.

Matthews, K.S., and Nichols, J.C. (1998). Lactose repressor protein: functional properties and structure. Prog. Nucleic Acid Res. Mol. Biol. *58*, 127–164.

McConnell, T.S., Cech, T.R., and Herschlag, D. (1993). Guanosine binding to the *Tetrahymena* ribozyme: thermodynamic coupling with oligonucleotide binding. Proc. Natl. Acad. Sci. USA *90*, 8362–8366.

McDaniel, B.A.M., Grundy, F.J., Artsimovitch, I., and Henkin, T.M. (2003). Transcription termination control of the S-box system: direct measurement of *S*-adenosylmethionine by the leader RNA. Proc. Natl. Acad. Sci. USA *100*, 3083–3088.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell *111*, 747–756.

Moszer, I., Glaser, P., and Danchin, A. (1995). SubtiList: a relational database for the *Bacillus subtilis* genome. Microbiology *141*, 261–268.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and

Breaker, R.R. (2002). Genetic control by a metabolite binding RNA. Chem. Biol. *9*, 1043–1049.

Ptashne, M., and Gann, A. (2002). Genes and Signals. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2002). Comparative genomics of thiamin biosynthesis in prokaryotes. J. Biol. Chem. *277*, 48949–48959.

Seetharaman, S., Zivarts, M., Sudarsan, N., and Breaker, R.R. (2001). Immobilized RNA switches for the analysis of complex chemical and biological mixtures. Nat. Biotechnol. *19*, 336–341.

Shu, D., and Guo, P. (2003). A viral RNA that binds ATP and contains a motif similar to an ATP-binding aptamer from SELEX. J. Biol. Chem. *278*, 7119–7125.

Soukup, G.A., and Breaker, R.R. (1999a). Engineering precision RNA switches. Proc. Natl. Acad. Sci. USA *96*, 3584–3589.

Soukup, G.A., and Breaker, R.R. (1999b). Relationship between internucleotide linkage geometry and the stability of RNA. RNA *5*, 1308–1325.

Stormo, G.D., and Ji, Y. (2001). Do mRNAs act as direct sensors of small molecules to control their expression? Proc. Natl. Acad. Sci. USA *98*, 9465–9467.

Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA *9*, 644–647.

Switzer, R.L., Zalkin, H., and Saxild, H.H. (2002). Purine, pyrimidine, and pyridine nucleotide metabolism. In *Bacillus subtilis* and Its Closest Relatives: From Genes to Cells. A.L. Sonenshein, *et al*., eds. (Washington: ASM Press), pp. 255-269.

Switzer, R.L., Turner, R.J., and Lu, Y. (1999). Regulation of the *Bacillus subtilis* pyrimidine biosynthetic operon by transcriptional attenuation: control of gene expression by an mRNA-binding protein. Prog. Nucleic Acid Res. Mol. Biol. *62*, 329–367.

Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2002). Regulation of riboflavin biosynthesis and transport genes in bacteria by translational and transcriptional attenuation. Nucleic Acids Res. *30*, 3141–3151.

Weng, M., Nagy, P.L., and Zalkin, H. (1995). Identification of the *Bacillus subtilis pur* operon repressor. Proc. Natl. Acad. Sci. USA *92*, 7455–7459.

White, H.B., III. (1976). Coenzymes as fossils of an earlier metabolic state. J. Mol. Evol. *7*, 101–104.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002a). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.

Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002b). An mRNA structure that controls gene expression by binding FMN. Proc. Natl. Acad. Sci. USA *99*, 15908–15913.